Quantization for Distributed Testing of Independence

Minna Chen, Wei Liu and Biao Chen
Syracuse University
Syracuse, NY, USA.
mchen18{wliu28,bichen}@syr.edu

John Matyjas
Air Force Research Laboratory
Rome, NY, USA.
matyjasj@rl.af.mi

Abstract – *We consider the problem of* distributed *test of statistical independence under communication constraints. While independence test is frequently encountered in various applications,* distributed *independence test is particularly useful for events detection in sensor networks: data correlation often occurs among sensor observations in the presence of a target. Focusing on the Gaussian case because of its tractability, we study in this paper the characteristics of optimal scalar quantizers for distributed test of independence where the optimality is in the sense of optimizing the error exponent. We also discuss the optimal quantizer properties for the finite sample regime, i.e., that of directly minimizing the error probability.*

Keywords: Distributed signal processing, test of independence, sensor networks.

1 Introduction

Test of statistical independence has been a classcal inference problem [1] and has found a wide range of applications, e.g., in image processing [2], economics [3]. The emerging wireless sensor networks bring new dimensions and challenges to this classical problem as the data are no longer centrally available. Dependence detection in distributed systems is often the first and crucial step in event detection/identification; thus its relevance in various sensor network applications is quite evident. One particular example, which will be used later is cooperative spectrum sensing in cognitive radio network: the presence of the primary user's signal introduces dependence among the decentralized spectrum sensors.

Take, for example, the Gaussian case and consider the following hypothesis testing problem: a pair of random sequences (X_i, Y_i), $i = 1, \cdots, n$, with (X_i, Y_i) independent and identically distributed (i.i.d.) according to the joint probability density function

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)\right).$$

The two hypothese under test are

$$\begin{cases} H_0: & \rho \neq 0 \\ H_1: & \rho = 0 \end{cases} \tag{1}$$

i.e., (X, Y) is bivariate Gaussian and they are independent under H_1 and dependent under H_0. Notice that assuming zero mean and unit variance does not lose any generality as long as the mean values and variances are known. In the centralized case where X and Y sequences are available, this statistical inference problem can be solved straightforwardly by applying some standard statistical inference frameworks depending on the situations (e.g., whether or not ρ is known under H_0) [4].

The problem becomes much more interesting and complicated when X and Y are not directly available; instead, compressed versions of X and Y subject to some rate constraints are used for the test of independence. This distributed test of independence is the focus of the present paper. To be more specific, we assume that X and Y are available respectively at two distributed sensor nodes. The sensor nodes communicate their data to the fusion center under a communication constraint of R_1 and R_2 bits per observation. The fusion center, upon receiving the sensor data, makes a final decision on whether X and Y are correlated or not. Our attempt is to understand properties of optimal quantizers at distributed nodes where the optimality is associated with the performance at the fusion center with regard to the dependence test.

Consider first the large sample regime, i.e., n is large. Given that (X_i, Y_i) form an i.i.d. sequence, it is easy to show that any reasonable quantizers will lead to a test with diminishing error probability as n grows for $R_1 > 0$ and $R_2 > 0$. Thus a sensible criterion is the speed with which the error probability approaches zero, i.e., the error exponent characterization. This is indeed the underlying reason for the problem setting where the null hypothesis H_0 represents dependence while independence occurs under H_1. Applying Stein's lemma [5]

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **JUL 2010**	2. REPORT TYPE	3. DATES COVERED **00-00-2010 to 00-00-2010**

4. TITLE AND SUBTITLE **Quantization for Distributed Testing of Independence**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Air Force Research Laboratory,Rome ,NY,13441**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Presented at the 13th International Conference on Information Fusion held in Edinburgh, UK on 26-29 July 2010. Sponsored in part by Office of Naval Research, Office of Naval Research Global, and U.S. Army Research Laboratory's Army Research Office (ARO).

14. ABSTRACT
We consider the problem of distributed test of statistical independence under communication constraints. While independence test is frequently encountered in various applications, distributed independence test is particularly useful for events detection in sensor networks: data correlation often occurs among sensor observations in the presence of a target. Focusing on the Gaussian case because of its tractability, we study in this paper the characteristics of optimal scalar quantizers for distributed test of independence where the optimality is in the sense of optimizing the error exponent. We also discuss the optimal quantizer properties for the finite sample regime, i.e., that of directly minimizing the error probability.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**5**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

to the hypothesis testing problem (1), for a given type I error constraint, the error exponent for the type II error (i.e., the Kullback Leibler distance between the distributions under H_0 and H_1) reduces to the mutual information between suitable random variables. For example, with centralized test, the optimal error exponent becomes $I(X;Y)$. Our focus in the large sample regime is to study quantizer properties in the context of distributed test against independence with Gaussian sources. Motivated by practical constraints that often require simple sensor processing, we consider only scalar quantizers at local sensors with 1 bit per observation. That is, $R_1 = R_2 = 1$ and each sensor quantizer is 'memoryless'. Our objective will be therefore to determine the optimal scalar quantizer structure that maximizes $I(U,V)$ where U and V are the one bit quantizer output for the two sensors.

Characterizing optimal error exponents for dependence test with communication constraints was first considered by Ahlswede and Csiszár [6]. In particular, for the special case of test of independence problem with one sided data compression, i.e., $R_2 = \infty$, a single letter characterization of the optimal error exponent was obtained in [6]. An overview of related work can be found in [7] and the references therein. We note here that the majority of the reported work are largely restricted to (X,Y) being discrete memoryless sources. Distributed test of independence with continuous alphabet sources (e.g., Gaussian sources) have been much less investigated.

We will also study distributed test of independence in the finite sample regime, that is, we attempt to characterize properties of quantizers that directly minimize error probability at the fusion center.

The rest of the paper is organized as follows. In Section 2, we give the problem statement and our main results. Section 3 are numerical examples. At last, we conclude in section 4.

2 Problem Statement and Main Results

2.1 Large sample regime

Consider the hypothesis test described in (1). The fusion center does not have direct access to the source sequence (X_i, Y_i), $i = 1, 2, \cdots, n$, but can be informed about the sources only at limited rates. Precisely, the local sensors apply scalar quantizers to their respective observations:

$$\begin{aligned} U_i &= \gamma_1(X_i) \\ V_i &= \gamma_2(Y_i) \end{aligned}$$

where U_i and $V_i \in \{0,1\}$.

For the large sample regime, the fusion center will decide H_0 or H_1 given the sequence (U_i, V_i) $i = 1, \cdots, n$ and we are to characterize the optimal quantizers that maximize the error exponent. Using the Neyman-Pearson criterion, we assume that the rejection region is the set $B \subset \mathcal{X}^n$ whose complement of B is $\bar{B}$. The minimum probability of type II error for a prescribed arbitrary small probability of type I error, denoted by $\beta_{R_1,R_2}(n,\epsilon)$, is defined as

$$\beta_{R_1,R_2}(n,\epsilon) = \min_B \{Q^n(\bar{B}) | B \subset \mathcal{X}^n, P^n(B) \leq \epsilon\} \quad (2)$$

The error exponent associated with $\beta_{R_1,R_2}(n,\epsilon)$ is, under the problem setup, the mutual information between U and V, $I(U,V)$. Our problem becomes finding a pair of binary quantizers such that $I(U,V)$ is maximized.

By restricting each sensor to a one bit scalar quantizer, we have the following result.

Theorem 1 *For the distributed test of independence problem described in (1) where each local quantizer is restricted to be one bit scalar quantizer with a single threshold, the optimal quantizers that maximize the error exponent are a sign detector, i.e., a binary quantizer with threshold*

$$t_1 = t_2 = 0 \quad (3)$$

While the result is rather intuitive with the symmetric problem setting, the proof is rather lengthy and is sketched in the Appendix. Notice that the result relies on the assumption of a single threshold quantizer: it is not known if such restriction may be relaxed though it appears to be the case from extensive numerical examples.

2.2 Finite sample regime

For the finite sample regime, we consider a Bayesian approach where the priors for the two hypotheses are assumed to be π_0 and π_1 respectively. We derive quantizer properties for minimum error probability with both two-sided and one-sided compression, with the latter refering to the situation in which the fusion center has full data from one sensor while compressed data from another. This situation arises naturally in the case where one of the sensors is tasked with the final decision making.

For the finite sample regime, we adopt the person-by-person optimal approach and obtain the following result for two-sided compression, following standard approach described in [8].

Proposition 1 *For the distributed testing of independence problem with one bit quantization defined above. If we further assume the fusion rule satisfies,*

$$\begin{aligned} P(U_0 = 1 | U = 1, V = j) &\geq P(U_0 = 1 | U = 0, V = j) \\ P(U_0 = 0 | U = 0, V = j) &\geq P(U_0 = 0 | U = 1, V = j) \end{aligned}$$

$$A_i = \sum_{j=0}^{1}[P(U_0=1|U_i=1,U_{\bar{i}}=j) - P(U_0=1|U_i=0,U_{\bar{i}}=j)]P(U_{\bar{i}}=j|x_{\bar{i}}) \quad (4)$$

$$B_i = \sum_{j=0}^{1}[P(U_0=0|U_i=0,U_{\bar{i}}=j) - P(U_0=0|U_i=1,U_{\bar{i}}=j)]P(U_{\bar{i}}=j|x_{\bar{i}}) \quad (5)$$

for all $j = \{0,1\}$, then the optimal local decision rule at ith sensor is given by:

$$P(U_i = 1|x_i) = \begin{cases} 1 & \text{if } \dfrac{\int_{x_{\bar{i}}} B_i P(x_{\bar{i}}|x_i H_1)dx_{\bar{i}}}{\int_{x_{\bar{i}}} A_i P(x_{\bar{i}}|x_i H_0)dx_{\bar{i}}} \geq \frac{\pi_0}{\pi_1} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where $\bar{i} = 3 - i$, for $i = 1, 2$, (hence, $x_{\bar{1}} = Y$), $\pi_0 = P(H_0)$, $\pi_1 = P(H_1)$, and A_i, B_i, $i = 1, 2$, defined in (4) and (5) at the top of next page.

If furthermore the fusion center uses the AND rule, we have

Proposition 2 *For the distributed test of independence problem with one bit quantization defined above, if we assume further that AND rule is used at the fusion center, i.e., $U_0 = 1$ if and only if $U = V = 1$, then the optimal local decision rule is given by:*

$$P(U_i = 1|x_i) = \begin{cases} 1 & \text{if } \dfrac{\int_{D_{\bar{i}}} P(x_{\bar{i}}|x_i H_1)dx_{\bar{i}}}{\int_{D_{\bar{i}}} P(x_{\bar{i}}|x_i H_0)dx_{\bar{i}}} \geq \frac{\pi_0}{\pi_1} \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

where $D_i = \{x_i : P(U_i = 1|x_i) = 1\}$ is the rejection region for hypothesis H_0 at ith local sensor.

For the case of one sided hypothesis testing of independence, e.g., $H_0 : \rho > 0$ versus $H_1 : \rho = 0$, we have the following corollary.

Corollary 1 *For the distributed one sided hypothesis testing of independence problem with one bit quantization defined above, single semi-infinite intervals for D_1 and D_2 form a PBPO solution for minimum probability of error.*

The fact that optimal quantizer has semi-infinite quantization intervals is rather appealing as it allows efficient search of a single threshold for quantizer design. Proof of Propositions 1 and 2 as well as Corollary 1 is omitted due to space limit.

3 Numerical examples

Fig. 1 plots $I(U;V)$ as a function of thresholds t_1 and t_2 for $\rho = 0.65$. Apparently $I(U;V)$ achieves its maximum (≈ 0.15) when $(t_1, t_2) = (0, 0)$. We further conjecture that, this point is actually a global maximum which is corroborated by extensive numerical results. The difficulty in proving it's global maximum is

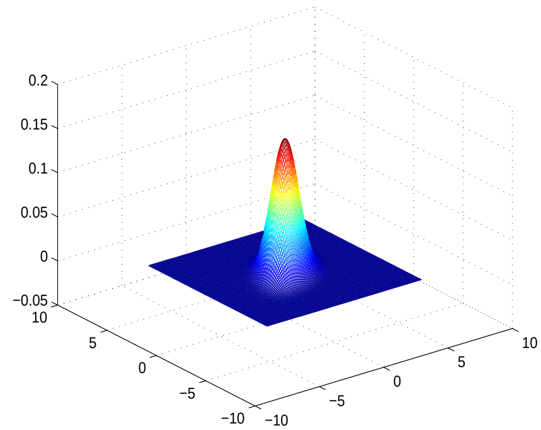

Figure 1: Plot of $I(U;V)$ as a function of thresholds t_1 and t_2 for $\rho = 0.65$.

that we do not have an analytical expression of the cumulative distribution function for a bivariate Gaussian distributed random variables.

An interesting application of our main result is the spectrum sensing problem in cognitive radio network, where multiple secondary users collaborate to detect whether the primary user is present or not. While the problem is well understood when the primary user's signal is fully observed (possibly corrupted by noise), it is more challenging when only a finite bits of information from each secondary user can be communicated to a decision maker. Consider the following simple model in which local sensors Y_1 and Y_2 receive a noisy version of the original signal through independent additive Gaussian channels.

$$Y_1^n = X^n + N_1^n, \quad (8)$$
$$Y_2^n = X^n + N_2^n, \quad (9)$$

where X^n is a n length samples of the primary user's signal, i.e., $X^n = [x_1, x_2, \cdots, x_n] \neq 0$ when the primary user is transmitting (hypothesis H_0) and $X^n = [0, 0, \cdots, 0]$ when the primary user is silent (hypothesis H_1). In this example, we assume that X is a zero mean independent Gaussian random process with variance P, for simplicity, which may be justified by the use of Gaussian pulse shaping filter used in digital communications. The noise N_1 and N_2 are independent standard Gaussian random variables.

Upon receiving Y_i^n, sensor i will send a binary decision vector u_i^n to the fusion center, the fusion center

will then decide wether the original signal is present or not. Clearly, if X is present, the received signals at local sensors Y_1 and Y_2 are correlated (In the simulation, we choose $P = 2.857$ to make sure that the correlation of Y_1 and Y_2 under H_0 is 0.65). We use the following decision rules, for $k = 1, 2$ and $i = 1, 2, \cdots, n$

$$u_{ki} = 1 \quad \text{if} \quad \frac{y_{ki}}{\sqrt{P+1}} > t \tag{10}$$

The fusion center decides $u_{0i} = 1$ if and only if $u_{1i} = u_{2i}$ for $i = 1, 2, \cdots, n$, and then make a final decision using the following majority rule,

$$u = 1 \quad \text{if} \quad \sum_{i=1}^{n} u_{0i} \geq t_0(n) \tag{11}$$

where $t_0(n)$ is chosen so that the probability of type I error $P_{e1} = 0.1$. Since, under H_0, $\sum_{i=1}^{n} u_{0i}$ is a binomial distribution with probability of success $p = Pr_0(u_{1i} = u_{2i})$, which can be easily calculated, $t_0(n)$ can be easily evaluated numerically for each n. Randomization is used to ensure that the false alarm probability to be precisely 0.1 so as to maximize the detection probability.

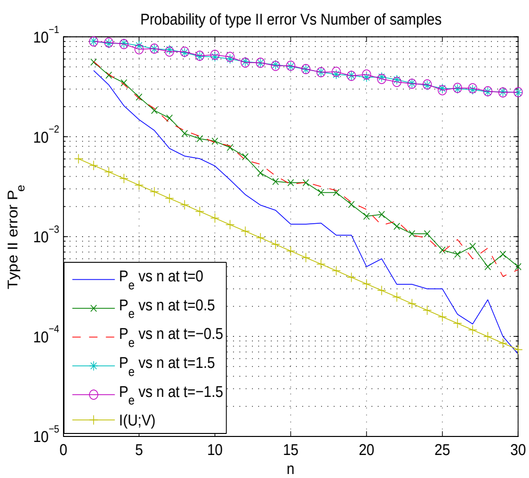

Figure 2: Probability of error for spectrum sensing.

Fig. 2 shows the performance of the above algorithm. In the simulation, we assume that $Pr(H_0) = 0.8$, and we choose five different local decision thresholds ($t = -1.5, -0.5, 0, 0.5, 1.5$) in (10). To compare the performance, we also plot the optimal error exponent $I(U;V)$ as plotted in Fig. 1. We observe that as number of samples increases, the probability of error decreases, and the threshold $t = 0$ performs the best among others. Notice that the vertical axis is in logarithmic scale and the slope appears to be equal to the plotted $I(U;V)$.

4 Conclusion

In this paper, we studied distributed test of independence of bivariate Gaussian sources with communication constraints. In particular, with one bit quantization, we derived quantization rules for single threshold quantizer at the local sensors that optimize the error exponent. For distributed one sided independence testwe proved that semi-infinite interval quantizers form a person by person optimal (PBPO) solution for minimum probability of error.

Appendix - Proof of Theorem 1

With one bit scalar quantization, optimizing error exponent is equivalent to maximize the mutual information $I(U;V)$. Define, under H_0, $P_{ij} = Pr(U = i; V = j), \quad i, j = \{0, 1\}$, which can be expressed in terms of integration of (1) given the single threshold quantizer assumption. By definition,

$$Pr(U = 1) = Pr(X \geq t_1) = Q(t_1) \tag{12}$$
$$Pr(V = 1) = Pr(Y \geq t_2) = Q(t_2) \tag{13}$$

where the Q function is complementary cumulative distribution function for standard Gaussian distribution. We want to maximize $I(U;V)$, where

$$\begin{aligned} I(U;V) &= H(U) + H(V) - H(U;V) && (14)\\ &= H(Q(t_1)) + H(Q(t_2)) \\ &\quad -H(P_{00}, P_{01}, P_{10}, P_{11}), && (15) \end{aligned}$$

where $H(\cdot)$ is the Shannon entropy function.

We now compute the first partial derivative of $I(U;V)$ with respect to t_1 and t_2, repectively. We get, with tedious but straightforward computation,

$$\begin{aligned} \frac{\partial I(U;V)}{\partial t_1} &= \frac{1}{\sqrt{2\pi}} \exp\frac{-t_1^2}{2} \{ \log\frac{Q(t_1)}{1-Q(t_1)} \\ &\quad +[1 - Q(\frac{t_2 - \rho t_1}{\sqrt{1-\rho^2}})] \log\frac{P_{00}}{P_{10}} \\ &\quad +Q(\frac{t_2 - \rho t_1}{\sqrt{1-\rho^2}}) \log\frac{P_{01}}{P_{11}} \} && (16)\\ \frac{\partial I(U;V)}{\partial t_2} &= \frac{1}{\sqrt{2\pi}} \exp\frac{-t_2^2}{2} \{ \log\frac{Q(t_2)}{1-Q(t_2)} \\ &\quad +[1 - Q(\frac{t_1 - \rho t_2}{\sqrt{1-\rho^2}})] \log\frac{P_{00}}{P_{01}} \\ &\quad +Q(\frac{t_1 - \rho t_2}{\sqrt{1-\rho^2}}) \log\frac{P_{10}}{P_{11}} \} && (17) \end{aligned}$$

One can easily check that $(t_1, t_2) = (0, 0)$ is a critical point, i.e., the first partial derivatives equal 0. We next check its Hessian matrix:

$$M = \begin{pmatrix} a(\rho) & b(\rho) \\ b(\rho) & c(\rho) \end{pmatrix} \tag{18}$$

where

$$a(\rho) = \frac{\partial^2 I(U;V)}{\partial t_1^2}|_{(t_1,t_2)=(0,0)}$$

$$b(\rho) = \frac{\partial^2 I(U;V)}{\partial t_1 t_2}|_{(t_1,t_2)=(0,0)}$$

$$c(\rho) = \frac{\partial^2 I(U;V)}{\partial t_2^2}|_{(t_1,t_2)=(0,0)}$$

We want to show that $a(\rho) < 0$ and $\det M = b(\rho)^2 - a(\rho)c(\rho) > 0$ for all $\rho \in [-1,0) \cup (0,1]$.

We can easily calculate that,

$$a(\rho) = c(\rho) \quad (19)$$

$$= \frac{1}{2\pi}[-4 + \frac{2\rho}{\sqrt{1-\rho^2}} \log \frac{P_{10}}{P_{11}} + \frac{1}{4P_{10}P_{11}}]|_{(0,0)} \quad (20)$$

$$b(\rho) = \frac{1}{2\pi}[\frac{2}{\sqrt{1-\rho^2}} \log \frac{P_{11}}{P_{10}} + \frac{P_{10}-P_{11}}{2P_{10}P_{11}}]|_{(0,0)} \quad (21)$$

Next, we introduce a lemma concerning evaluating the cumulative distribution function of a standard bivariate Gaussian distribution at point $(0,0)$.

Lemma 1 *[9, page 290]*

$$P_{00}(t_1 = t_2 = 0) = P_{11}(t_1 = t_2 = 0) = \frac{1}{4} + \frac{1}{2\pi}\arcsin(\rho) \quad (22)$$

$$P_{01}(t_1 = t_2 = 0) = P_{10}(t_1 = t_2 = 0) = \frac{1}{4} - \frac{1}{2\pi}\arcsin(\rho) \quad (23)$$

Using (22) and (23), we can further get

$$a(\rho) = c(\rho) = \frac{1}{2\pi}[-4 + \frac{2\rho}{\sqrt{1-\rho^2}} \log \frac{\pi - 2\arcsin\rho}{\pi + 2\arcsin\rho} + \frac{4\pi^2}{\pi^2 - 4\arcsin^2\rho}] \quad (24)$$

$$b(\rho) = \frac{1}{2\pi}[\frac{2}{\sqrt{1-\rho^2}} \log \frac{\pi + 2\arcsin\rho}{\pi - 2\arcsin\rho} - \frac{8\pi\arcsin\rho}{\pi^2 - 4\arcsin^2\rho}] \quad (25)$$

Next, we want to evaluate functions $a(\rho), b(\rho)$ and $c(\rho)$ with the help of the following two lemmas.

Lemma 2 *For $a(\rho)$ and $c(\rho)$ defined above, we have:*

$$a(\rho) = c(\rho) \leq 0 \quad (26)$$

for all $\rho \in [-1,1]$ and the maximum is achieved when $\rho = 0$.

Lemma 3 *For the function $b(\rho)$ defined above, we have:*

$$b(\rho) > 0, \quad \textit{if} \quad \rho \in (0,1] \quad (27)$$

$$b(\rho) < 0, \quad \textit{if} \quad \rho \in [-1,0) \quad (28)$$

$$b(\rho) = 0, \quad \textit{if} \quad \rho = 0 \quad (29)$$

Form Lemma 2, we can see that $a(\rho) < 0$ for all $\rho \in [-1,0) \cup (0,1]$ is satisfied. Next, we want to prove that $b^2(\rho) - a(\rho)c(\rho) < 0$ for all $\rho \neq 0$ is also true. Notice that, form Lemmas 2 and 3, we only need to prove that

$$-a(\rho) > b(\rho) \quad \text{if} \quad \rho \in (0,1] \quad (30)$$

$$a(\rho) < b(\rho) \quad \text{if} \quad \rho \in [-1,0) \quad (31)$$

Define, $d(\rho) = -a(\rho) - b(\rho)$ and $e(\rho) = a(\rho) - b(\rho)$. We want to show that

$$d(\rho) > 0 \quad \text{if} \quad \rho \in (0,1] \quad (32)$$

$$e(\rho) < 0 \quad \text{if} \quad \rho \in [-1,0) \quad (33)$$

This can be verified by noting that

$$d(\rho) = \frac{1}{2\pi}[-2\sqrt{\frac{1-\rho}{1+\rho}} \log \frac{\pi + 2\arcsin\rho}{\pi - 2\arcsin\rho} + \frac{8(\pi - 2\arcsin\rho)\arcsin\rho}{\pi^2 - 4\arcsin^2\rho}] \quad (34)$$

$$e(\rho) = \frac{1}{2\pi}[2\sqrt{\frac{1+\rho}{1-\rho}} \log \frac{\pi - 2\arcsin\rho}{\pi + 2\arcsin\rho} + \frac{8(\pi + 2\arcsin\rho)\arcsin\rho}{\pi^2 - 4\arcsin^2\rho}] \quad (35)$$

References

[1] W. Hoeffding, "A non-parametric test of independence," *The Annals of Mathematical Statistics*, vol. 19, no. 4, pp. 546–557, 1948.

[2] J. G. Daugman, "High confidence visual recognition of persons by a test ofstatistical independence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1148–1161, Nov 1993.

[3] W. A. Broock, J. A. Scheinkman, W. D. Dechert, and B. LeBaron, "A test for independence based on the correlation dimension," *Econometric Reviews*, vol. 15, no. 3, pp. 197–235, 1996.

[4] D.R. Cox and D.V. Hinkley, *Theoretical Statistics*, Chapman and Hall Ltd, New York, 1974.

[5] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[6] R. Ahlswede and I. Csiszar, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, no. 4, pp. 533–542, July 1986.

[7] T. S. Han and S. I. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2300–2324, Oct. 1998.

[8] P. K. Varshney, *Distributed Detection and Data Fusion*, Springer-Verlag New York, Inc, New York, NY, 1997.

[9] H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, 1946.